

## **Exhibit G**

Methodology article

Open Access

## SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones

Ruth Wellenreuther\*, Ingo Schupp, The German cDNA Consortium, Annemarie Poustka and Stefan Wiemann

Address: Department of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany

Email: Ruth Wellenreuther\* - [r.wellenreuther@dkfz.de](mailto:r.wellenreuther@dkfz.de); Ingo Schupp - [i.schupp@dkfz.de](mailto:i.schupp@dkfz.de); The German cDNA Consortium - [s.wiemann@dkfz.de](mailto:s.wiemann@dkfz.de); Annemarie Poustka - [a.poustka@dkfz.de](mailto:a.poustka@dkfz.de); Stefan Wiemann - [s.wiemann@dkfz.de](mailto:s.wiemann@dkfz.de)

\* Corresponding author

Published: 15 June 2004

BMC Genomics 2004, 5:36

This article is available from: <http://www.biomedcentral.com/1471-2164/5/36>

Received: 05 March 2004

Accepted: 15 June 2004

© 2004 Wellenreuther et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** cDNA libraries are widely used to identify genes and splice variants, and as a physical resource for full-length clones. Conventionally-generated cDNA libraries contain a high percentage of 5'-truncated clones. Current library construction methods that enrich for full-length mRNA are laborious, and involve several enzymatic steps performed on mRNA, which renders them sensitive to RNA degradation. The SMART technique for full-length enrichment is robust but results in limited cDNA insert size of the library.

**Results:** We describe a method to construct SMART full-length enriched cDNA libraries with large insert sizes. Sub-libraries were generated from size-fractionated cDNA with an average insert size of up to seven kb. The percentage of full-length clones was calculated for different size ranges from BLAST results of over 12,000 5'ESTs.

**Conclusions:** The presented technique is suitable to generate full-length enriched cDNA libraries with large average insert sizes in a straightforward and robust way. The representation of full-coding clones is high also for large cDNAs (70%, 4–10 kb), when high-quality starting mRNA is used.

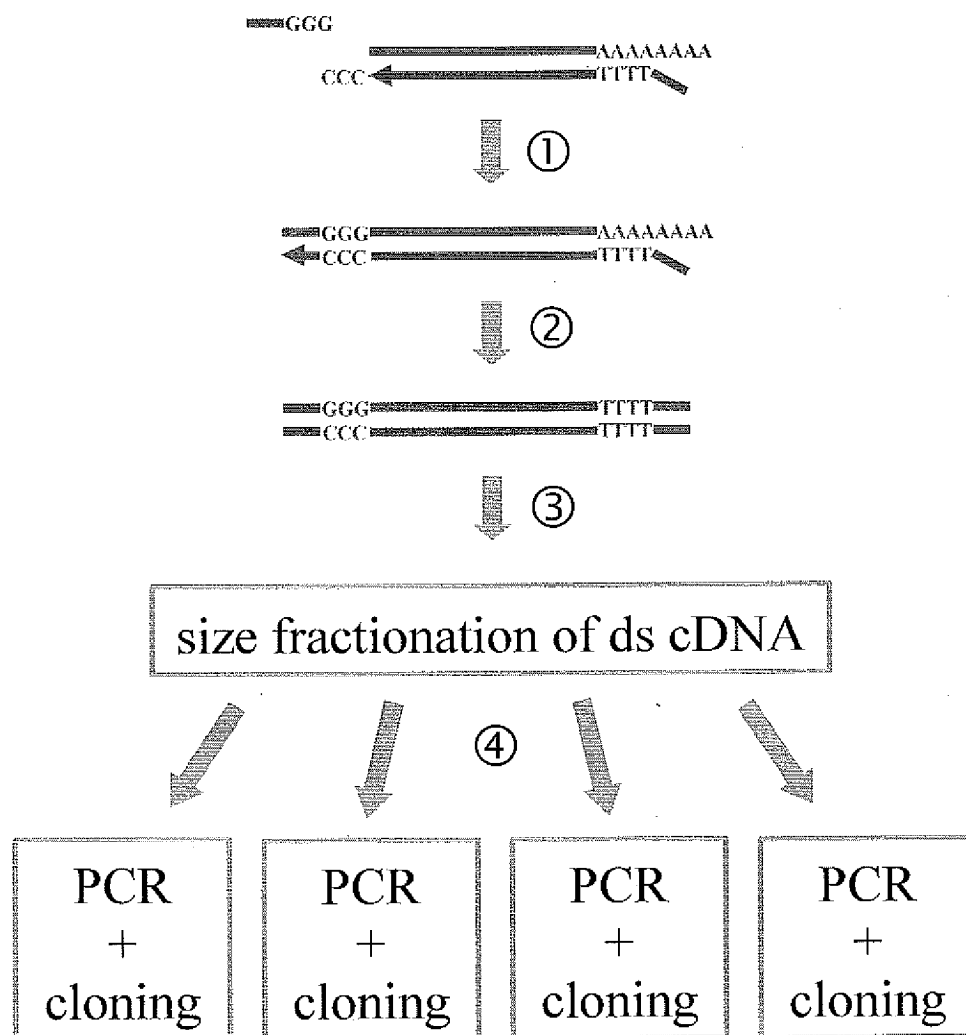
### Background

Full-length cDNA clones are indispensable tools for functional genomics [1]. cDNA libraries are widely used to identify genes and splice variants and as a physical resource for full-length clones. Unfortunately, cDNA libraries constructed according to conventional methods [2] contain a high percentage of 5' truncated clones due to the premature stop of reverse transcription (RT) of the template mRNA. This is especially true for large mRNAs and those tending to form secondary structures. In addition, there is a size bias against large fragments inherent in the cloning procedure. For these reasons, large full-length

cDNAs are strongly underrepresented in conventional libraries. Several methods have been developed to construct cDNA libraries that are enriched for full-length cDNAs. Most are based on either RNA oligo ligation to the 5' end of mRNA [3,4], 5' cap affinity selection via eukaryotic initiation factor 4E [5], or 5' cap biotinylation followed by biotin affinity selection [6,7]. Common to these methods is that they are laborious and contain several enzymatic steps that must be performed on mRNA. Therefore, they are sensitive to quality loss through RNA degradation. Furthermore, they require high amounts of starting mRNA (5–100 µg depending on method).

In contrast, using SMART technology for full-length enrichment of cDNA is very straightforward and robust and requires only 0.025–1 µg of starting mRNA [8]. This technology utilizes the property of some MMLV reverse transcriptases to add a few C residues at the 3' end of the first strand cDNA when they reach the end of the mRNA template, but not at prematurely terminated reverse transcripts. An RNA oligo ending in three G residues and present in the reaction, forms base pairings with the added Cs and serves as a prolonged template for reverse transcription. By these means, full-length cDNAs but not prematurely terminated ones are 5'-tagged and can be amplified by an RNA oligo-specific primer (figure 1, step 1 and 2). The percentage of full-length clones in libraries

constructed with the SMART technique is much higher compared to conventional libraries [8] and, when transcripts up to 3 kb are compared, better than libraries constructed with other full-length enriching techniques [9]. However, large clones are rarely found in SMART libraries as well as in libraries constructed according to the other full-length enriching techniques [8,9], unless specialized lambda vectors are used [10]. We modified the proven and robust SMART technique to construct cDNA libraries with large average insert sizes in convenient plasmid vectors. We here report the construction of size fraction sub-libraries enriched for full-length clones having an average insert size of up to 7 kb and the analysis of full-length percentage for these libraries.



**Figure 1**

**Library construction process** Schematic representation of the library construction process. First strand cDNA is synthesized with the SMART system (1), second strand synthesis is primed by a SMART oligo-specific primer (2), double-stranded cDNA is size-fractionated via agarose gel electrophoresis (3), and size fractions are amplified and cloned separately (4).

## Results and Discussion

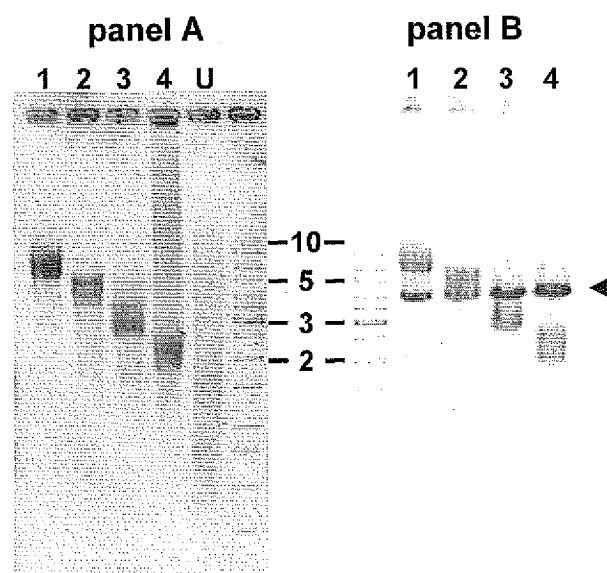
### Generation of size-fractionated full-length enriched cDNA

cDNA synthesis according to the SMART protocol is as convenient as conventional first strand synthesis. There is no need for any mRNA manipulative step prior to the reaction and the only difference is the presence of the SMART RNA oligo in the reaction (figure 1, step 1 and 2). Synthesis must be done with a MMLV RTase that is RNaseH negative to ensure addition of C residues, and to prevent SMART RNA oligo degradation during base pairing with these residues. The full-length selective step is the PCR amplification following cDNA synthesis, therefore, there is a bias against large cDNAs, as smaller cDNAs are preferentially amplified during PCR (personal observation). In our strategy, cDNA is size-fractionated prior to this PCR amplification step and PCR is performed with the different size fractions in separate reactions (figure 1, step 3 and 4). Because large cDNAs are less frequent, more PCR cycles must be done on the large fractions compared to smaller fractions to obtain an equivalent amount of PCR product for cloning. But, to avoid increasing redundancy and to reduce errors introduced by PCR polymerase, as few PCR cycles as possible must be performed. We typically did 12 to 16 cycles, depending on size fraction. By these means, large cDNAs could be amplified as efficiently as smaller ones. These large cDNAs are strongly underrepresented in control amplification products of unfractionated cDNA (figure 2, panel A).

Polymerase error rate is a major concern in PCR-based library construction techniques. Therefore, it is crucial to perform as few PCR cycles as possible, as each duplication increases the number of introduced errors by a factor of two, assuming a constant error rate of the used polymerase. The Expand™ PCR System we used was tested to have an error rate of  $8.5 \times 10^{-6}$  [11]. Starting with PolyA+ RNA, we could restrict the number of cycles to 12 to 16. Levesque et al., who also combined SMART cDNA amplification with size fractionation, started with total RNA and did 45 to 47 cycles in total. In contrast to our approach, where amplification follows size fractionation, they did 33 cycles before and 12–14 after fractionation. In their study, the obtained sub-libraries were not analysed for insert size range, instead, they screened them with three gene-specific probes [12].

### Insert size of libraries

In conventionally-constructed libraries, large insert clones are rarely found. This is because very long transcripts often get truncated during cDNA synthesis, and because there is a strong size bias against large fragments inherent in the cloning procedure, i.e. ligation and bacterial transformation. In our strategy, PCR-amplified cDNA size fractions are restriction digested and separately cloned into a plasmid vector to obtain size fraction sub-libraries. To analyse



**Figure 2**

**Quality control of cDNA and sub-libraries** Panel A: Amplified cDNA size fractions. cDNA was size-fractionated and separate size fractions (1–4) were PCR-amplified. U = unfractionated control, size marker in kb. Panel B: Insert analysis of sub-libraries obtained by cloning of the amplified size fractions shown in panel A. Plasmid DNA of clone pools of 5,000–10,000 clones was restriction digested by Sfi I. An arrow head indicates the vector band. The smear corresponds to the insert size range of the the sub-library.

the range of insert sizes within these sub-libraries, clone pools of 5000–10,000 clones were grown in semi-solid agar and plasmid restriction digests of the clone pools were performed. Each sub-library almost exclusively contains inserts within the size range of the corresponding cDNA size fraction that was cloned to produce this sub-library (figure 2, panels A and B). In sub-library 1 for example, most inserts are between 6 and 8 kb. Such inserts are rarely found in conventional libraries.

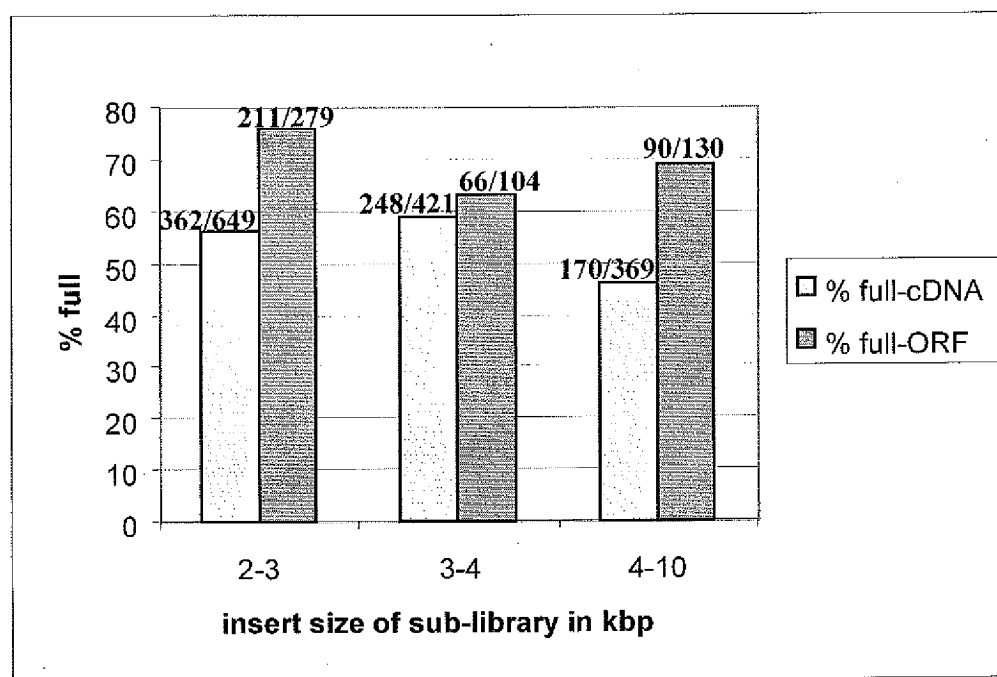
### Full-length content

The full-length enriched cDNA sub-libraries generated according to the protocol described here serve as clone resource for the cDNA sequencing efforts of the German cDNA Consortium <http://www.dkfz-heidelberg.de/mga/groups.asp?siteID=48> [13]. Within this project, over 100,000 5'ESTs have been generated. All sequences are submitted to public databases and clones are available through the German Resource Center for Genome Research <http://www.rzpd.de>. To determine the full-length cDNA content of our libraries, 5'ESTs were blasted against human RefSeq sequences according to parameters

specified in the Methods section. The total number of hits to known mRNAs were set as 100% and the percentage of clones containing the 5' end of the hit was calculated. Accordingly, full-ORF content was determined by BLAST analysis against the SWISSPROT database.

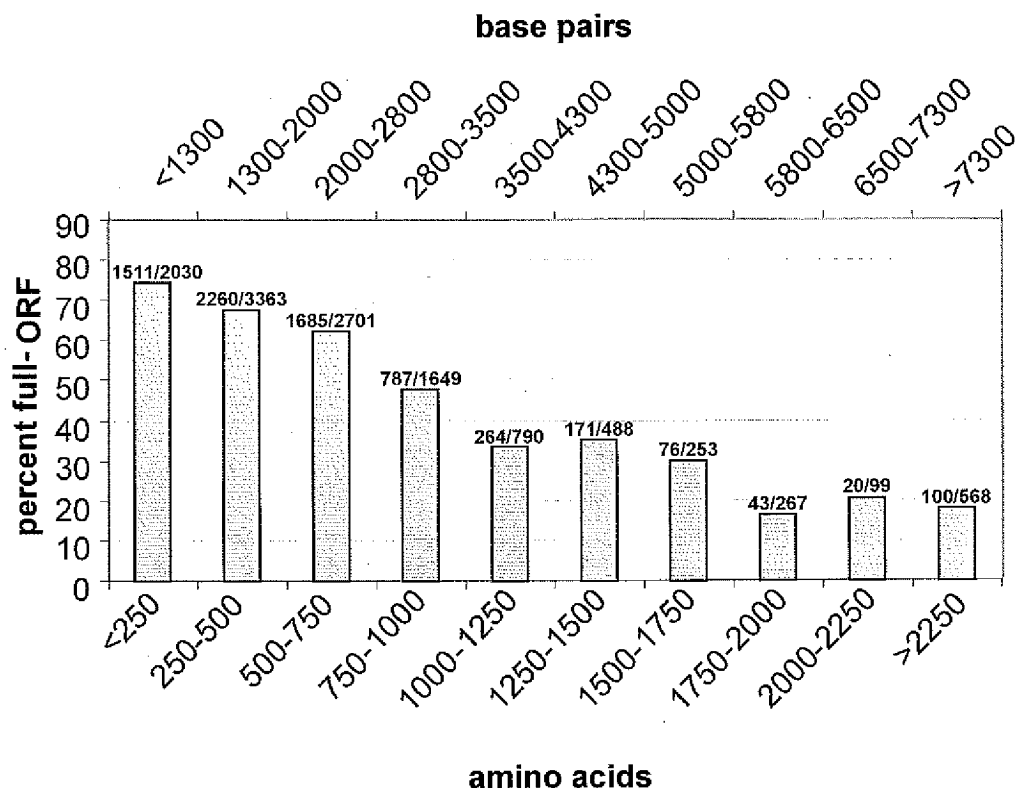
Figure 3 shows full-cDNA and full-ORF percentages of three size fraction sub-libraries made from an endometrium carcinoma cell line. With 3827 5'ESTs blasted in total for these three sub-libraries, 1439 hits to known mRNAs were found in human RefSeq and 513 hits to known proteins were found in SWISSPROT. For calculation of full-cDNA/full-ORF percentage, only hits within the corresponding size range of the sub-library were taken into consideration. Full-length percentages range from 46% to 59%, and full-ORF percentages from 63% to 76%, depending on sub-library. Full-length content does not decrease significantly with increasing cDNA/ORF size, as it is observed in conventional cDNA libraries. In the sub-library containing 4–10 kb inserts, the percentage of full-coding clones is still almost 70%, which is extremely high for this size range (figure 3).

Unlike other full-length enriching protocols, there is no negative selection against truncated mRNA molecules in the SMART technique, because the basic principle is selection for full-reverse transcribed mRNA molecules rather than mRNA cap selection. Therefore, mRNA quality is crucial. The starting mRNA of the library, which full-length analysis is shown in figure 3, was of highest quality, i.e. in a control agarose gel mRNA smear up to 10 kb. Figure 4 shows the analysis of sequence data from over ten different libraries made from mRNAs of various qualities. With 50,023 5'ESTs blasted in total, 12,208 hits to known proteins were found in the SWISSPROT database. Size windows shown correspond to the size of the SWISSPROT hits. For better orientation, the calculated corresponding mRNA/cDNA size is also shown. Here, full-ORF content decreases with increasing mRNA size, as can be expected due to the fact that for large transcripts there is a higher probability of truncated molecules in the starting mRNA. For transcripts up to 3 kb 60–70% contain the complete ORF. This number gradually decreases to 30% for 5–6 kb and 20% for 7 kb. Although these numbers are still reasonably high bearing transcript size in mind, they are lower than in the library shown in figure 3. Probably, this is due to lower quality of starting mRNAs.



**Figure 3**

**Completeness of cDNA and coding region in different size fraction sub-libraries of a cDNA library** Sequences from an endometrium carcinoma cell line library were analysed. Completeness of cDNA/ORF was calculated by BLAST analysis against human RefSeq ("full-length") / SWISSPROT ("full-ORF") database, respectively, with parameters specified in the Methods section. 3827 5'ESTs were blasted, 1439 hits were found in human RefSeq, and 513 hits in SWISSPROT. Percentages of full-length/full-ORF hits were calculated.

**Figure 4**

**Completeness of coding region in different size ranges** Completeness of ORF was calculated by BLAST against the SWISSPROT database, with parameters specified in the Methods section. The lower X-axis indicates the length of the SWISSPROT hit in amino acids. The upper X-axis indicates calculated corresponding cDNA lengths, assuming an average 3' UTR length of 525 bp [21].

cDNA size fractionation has been used previously in two studies to enrich cDNA libraries for full-length clones [12,14]. In both studies, the sub-libraries were not analysed for insert sizes. In consequence, it remains unclear, if the sub-libraries actually contained the expected range of insert sizes. Levesque et al. [12] also combined the SMART technique with cDNA size fractionation, but did not analyse the overall full-length content, instead, they screened the libraries with three gene-specific probes. Draper et al. [14] calculated the percentage of full-coding clones in size fractionated libraries from BLAST results of 78 hits in total and down to 3 hits per size range. We calculated the percentage of full-coding clones in the libraries generated according to the presented method from BLAST results of over 12,000 hits in total and between 99 and 3363 hits per size range. The high number of hits for a given size range permit a much more reliable calculation of full-length percentages compared to former studies. Furthermore, because of the large insert size of our sub-libraries, large size ranges can be analysed

(up to 10 kb), which had not been analysed before in similar studies [8,9,14].

### Conclusions

The method presented is attractive for the construction of full-length enriched cDNA libraries with large average insert sizes for several reasons. First, there is no additional enzymatic step for the enrichment, which saves time. Second, it is easy-to-use, as enzymatic steps performed on mRNA, which are necessary in other full-length enriching techniques, are extremely critical in terms of mRNA degradation and quantity loss. Third, the cDNA sizing protocol presented is very efficient and can be performed with basic laboratory equipment. cDNA libraries constructed according to the method presented also yield high full-length percentages for large cDNAs/ORFs when high quality starting mRNA is used.

## Methods

### cDNA synthesis

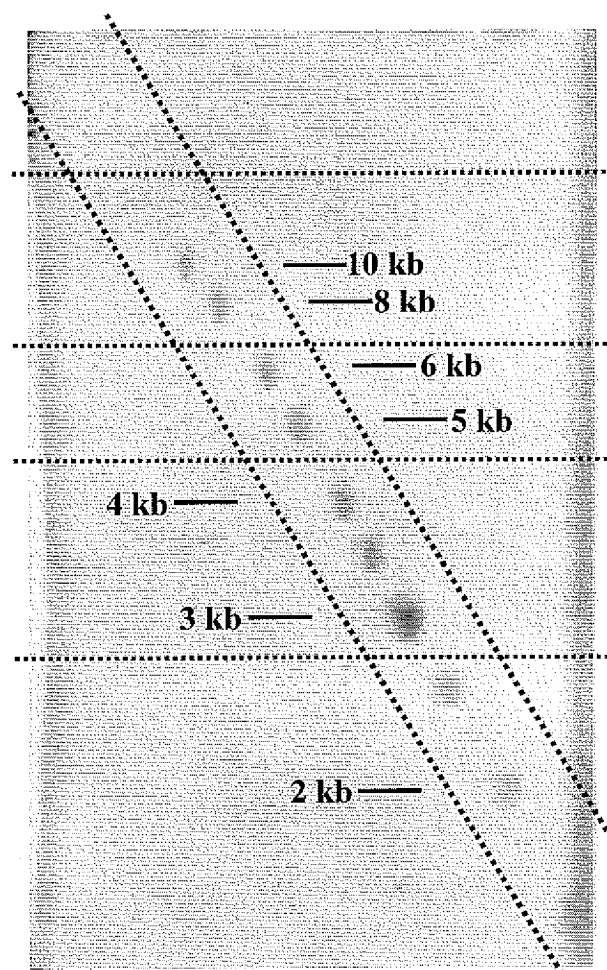
First strand cDNA was synthesized from 1 µg of mRNA with the "SMART cDNA Library Construction Kit" (Clontech) in a 10 µl reaction according to the manufacturers protocol. In this reaction, a fraction of full-transcribed first strand cDNA molecules but not truncated cDNAs is tagged with a short sequence complementary to the SMART oligo. The SMART oligo sequence (AAGCACTGG-TATCAACGCAGAGTGGCCATTACGGCCGGG) and the overhang of the oligo(dT) primer (ATTCTAGAGGCCGAGGCCGCCGACATG [dT]<sub>30</sub>VN) used for first strand synthesis both include a SfiI restriction site. After first strand synthesis, 40 pmol of 5' PCR primer (corresponding to the SMART oligo sequence) was added and first strand cDNA was denatured for 5 min at 95°C. The reaction was cooled to 60°C and second strand reaction mix was added to give a final concentration of 1x PCR reaction buffer (Expand 20 kb<sup>PLUS</sup> PCR System, Roche), 0.5 mM dNTPs, and 8.3 U/µl Expand 20 kb<sup>PLUS</sup> enzyme mix (Roche) in a volume of 60 µl. This second strand reaction mixture was incubated for 3 cycles of 15 min 60°C and 15 min 68°C. The second strand reaction was phenol-extracted and cDNA was precipitated from the aqueous phase with 1/2 volume 7.5 M ammonium acetate and 2.5 volumes of 100% ethanol. The washed pellet was dried and suspended in 10 µl of water. As a quality control, 1 µl was electrophoresed on an 1% agarose gel.

### Size fractionation of cDNA

Double-stranded cDNA alongside with a DNA ladder was subjected to size fractionation on two consecutive agarose gels: cDNA was separated on 0.7 % agarose gel and the size fraction 1.5–20 kb excised from the cDNA lane and from the DNA ladder lane. The gel slices were then rotated by 90°C and placed in a gel tray. A 1.4 % low-melting agarose gel was cast around the slices. Electrophoresis was performed over night at 37 V and the DNA ladder "lane" was stained and photographed with a ruler. 3–6 size fractions were excised from the unstained cDNA "lane" according to the DNA ladder "lane" (figure 5). cDNA was extracted from the gel slices with agarase (gelase, Epicentre) according to the manufacturers instructions. After gelase digestion, the reaction was phenol extracted, the aqueous phase incubated on ice for 15 min, and centrifuged at 4°C with maximum speed for 15 min. cDNA was precipitated from the supernatant with 1/2 volume 7.5 M ammonium acetate, 1 µl PelletPaint (Novagen), and 2.5 volumes of 100% ethanol. Washed pellets were dried and suspended in 10 µl of water.

### PCR amplification of cDNA size fractions

One µl of each cDNA fraction was amplified in a 10 µl reaction containing a final concentration of 1x PCR reaction buffer (Expand 20 kb<sup>PLUS</sup> PCR System, Roche), 0.5



**Figure 5**  
**cDNA size fractionation** Alongside with the cDNA, a DNA ladder is size-fractionated and stained. The ladder is photographed with a ruler and cutting edges are marked (thin dotted lines). The unstained cDNA is cut from the gel accordingly.

mM dNTPs, 0.5 pmol/µl forward primer (AAGCACTGG-TATCAACGCAGACT), 0.5 pmol/µl reverse primer (ATTCTAGAGGCCGAGGCCGCCGACATG), and 8.3 U/µl Expand 20 kb<sup>PLUS</sup> enzyme mix (Roche). To perform manual hot start, the reactions were prepared in two master mixes, one containing buffer and enzyme, the other containing dNTPs, primer, and cDNA. The two master mixes were combined at 92°C. After initial denaturation at 92°C for 3 min, 12–16 cycles (depending on size fraction and second strand cDNA quality and intensity) of 92°C 10 sec and 68°C 14 min were performed. PCR products were analysed on agarose gel and PCR was repeated

in a 50 µl volume with 5 µl cDNA and fine-tuned cycle number (i.e. reduced for intensive products and increased for weak signals). Five µl of 50 were analysed on an agarose gel. The remaining reaction was proteinase K digested, phenol extracted, and precipitated.

#### Cloning and quality control of sub-libraries

The precipitated amplified cDNA was SfiI-digested in a 40 µl volume. The SfiI digest was gel-purified using low-melting agarose and gelase (Epicentre). DNA was suspended in 10 µl water and concentration was determined using the PicoGreen reagent (Molecular Probes). 20 fmol of cDNA was ligated to 10 fmol Sfi-digested pSPORT1\_Sfi vector (a modified pSPORT vector having the part of the MCS between KpnI and HindIII exchanged by the corresponding part of the pTriplex2 MCS, so that it contains SfiI sites). For quality control, 5,000–10,000 clones were grown in semi-solid agar (SeaPrep agarose, BMA), centrifuged, plasmid DNA was extracted from these clone pools, SfiI-digested, and analyzed on an agarose gel. If the quality was satisfactory, 96 single clones were picked and insert analysis was performed as with the clone pools.

#### Examination of full-length clone content

Libraries were arrayed in 384-well plates and clones were randomly sequenced from the 5' end. 5'ESTs longer than 150 bp were compared to public databases using the BLAST algorithm [15,16] within the Heidelberg Unix Sequence Analysis Resources (HUSAR; <http://genome.dkfz-heidelberg.de/>) [17].

5' ESTs were compared to a human subset of RefSeq [18] by BLAST (default parameters, except a wordsize of 20 bp was used) to calculate the percentage of full-length cDNA clones. The BLAST outputs were further analysed with the following criteria to find the maximum scoring RefSeq entry: Minimum HSP length of 50 bp, start of HSP within the first 100 bp of 5'EST, end of HSP within the last 15% of 5'EST length, sequence identity within HSP at least 95%. If several HSPs within the same hit fit these criteria, the more upstream match was chosen. A clone was defined as "full length", when the 5'end of the 5'EST was upstream or up to 50 bp downstream of the start of the corresponding RefSeq entry. This last criteria was chosen to take into account the fact that transcription start site is variable for most genes [19], or even unknown.

To calculate the percentage of full-ORF clones, a BLAST search of the 5'ESTs against the SWISSPROT database [20] was performed with default parameters. HSPs with a length less than 20 amino acids and sequence identity below 75% were filtered out. A clone was calculated as full-ORF, when the most upstream HSP of the maximum scoring hit contained the first amino acid of the SWISS-PROT entry.

#### Abbreviations

SMART = Switching Mechanism At 5' end of RNA

UTR = untranslated region

ORF = open reading frame

RT = reverse transcription

HSP = High-scoring Segment Pair

#### Authors' contributions

cDNA library construction and development of the improved method was done by RW, 5'ESTs were generated within the German cDNA Consortium, BLAST analysis was done by IS; SW and AP initiated the project. All authors read and approved the final manuscript.

#### Acknowledgements

We thank Daniela Heiss and Nina Claudino for excellent technical assistance, clone picking, and management of the libraries, and Patricia McCabe for critical reading of the manuscript.

#### References

1. Wiemann S, Mehrle A, Bechtel S, Wellenreuther R, Pepperkok R, Poustka A: **CDNAs for functional genomics and proteomics: the German Consortium.** *C R Biol* 2003, **326**:1003-1009.
2. Gubler U, Hoffman BJ: **A simple and very efficient method for generating cDNA libraries.** *Gene* 1983, **25**:263-269.
3. Kato S, Sekine S, Oh SW, Kim NS, Umezawa Y, Abe N, Yokoyama-Kobayashi M, Aoki T: **Construction of a human full-length cDNA bank.** *Gene* 1994, **150**:243-250.
4. Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, Sugano S: **Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library.** *Gene* 1997, **200**:149-156.
5. Edery I, Chu LL, Sonenberg N, Pelletier J: **An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture).** *Mol Cell Biol* 1995, **15**:3363-3371.
6. Carninci P, Hayashizaki Y: **High-efficiency full-length cDNA cloning.** *Methods Enzymol* 1999, **303**:19-44.
7. Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M, Muramatsu M, Hayashizaki Y, Schneider C: **High-efficiency full-length cDNA cloning by biotinylated CAP trapper.** *Genomics* 1996, **37**:327-336.
8. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD: **Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction.** *Biotechniques* 2001, **30**:892-897.
9. Sugahara Y, Carninci P, Itoh M, Shibata K, Konno H, Endo T, Muramatsu M, Hayashizaki Y: **Comparative evaluation of 5'-end-sequence quality of clones in CAP trapper and other full-length-cDNA libraries.** *Gene* 2001, **263**:93-102.
10. Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraki T, Hirozane T, Watahiki A, Shibata K, Konno H, Muramatsu M, Hayashizaki Y: **Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis.** *Genomics* 2001, **77**:79-90.
11. Frey B, Suppmann B: **Demonstration of the Expand PCR System's Greater Fidelity and Higher Yield with a lacI-based PCR Fidelity Assay.** *Biochemica* 1995, **2**:8-9.
12. Levesque V, Fayad T, Ndiaye K, Nahe Diouf M, Lussler JG: **Size-selection of cDNA libraries for the cloning of cDNAs after suppression subtractive hybridization.** *Biotechniques* 2003, **35**:72-78.
13. Wiemann S, Well B, Wellenreuther R, Gassenhuber J, Glassl S, Ansorge W, Bocher M, Blocker H, Bauersachs S, Blum H, Lauber J, Dusterhoft A, Beyer A, Kohrer K, Strack N, Mewes HW, Ottenwal-



- der B, Obermaier B, Tampe J, Heubner D, Wambutt R, Korn B, Klein M, Poustka A: **Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs.** *Genome Res* 2001, **11**:422-435.
14. Draper MP, August PR, Connolly T, Packard B, Call KM: **Efficient cloning of full-length cDNAs based on cDNA size fractionation.** *Genomics* 2002, **79**:603-607.
  15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
  16. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
  17. Senger M, Flores T, Glatting K, Ernst P, Hotz-Wagenblatt A, Suhai S: **W2H: WWW interface to the GCG sequence analysis package.** *Bioinformatics* 1998, **14**:452-457.
  18. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence project: update and current status.** *Nucleic Acids Res* 2003, **31**:34-37.
  19. Suzuki Y, Taira H, Tsunoda T, Mizushima-Sugano J, Sese J, Hata H, Ota T, Isogai T, Tanaka T, Morishita S, Okubo K, Sakaki Y, Nakamura Y, Suyama A, Sugano S: **Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites.** *EMBO Rep* 2001, **2**:388-393.
  20. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M: **The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.** *Nucleic Acids Res* 2003, **31**:365-370.
  21. Pesole G, Liuni S, Grillo G, Licciulli F, Mignone F, Gissi C, Saccone C: **UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002.** *Nucleic Acids Res* 2002, **30**:335-340.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)



**BioMed Central**